Lustre + Linux

Convergence
Upstream client progress since last LUG

• More checkpatch and style cleanups
• Kept LNet/libcfs in sync with pre-multirail code
• Lustre updated from 2.5 to pre-2.9 version
• Some sites are using the upstream client in production
• Clearer roadmap to leave staging
• Initial review by the kernel maintainers
Impact of a functional Upstream client

- Perks from upstream for OpenSFS/Intel branch
  - Initial support of newer kernels
  - Master can work with latest [M]OFED stacks
- Real bug fixes flowing both ways
- Preparing automated testing of upstream client
  - Can now test latest staging-testing branch.
  - Working out test failures (LU-4011)
  - Don't test latest functionality (LU-7344)
  - Goal: automate testing from email submitted patches
Upstream backports to OpenSFS/Intel branch

- Patches for upstream not always reviewed + tested.
  - Push work to OpenSFS branch to review
  - Finding bugs. Fixes pushed upstream.

- Backport support up to 4.9+ kernels
  - Goal to match latest kernel for 2.10 release

- Backport of sysfs/debugfs
  - Need to make lctl set_param -P really work

- Use 64 bit kernel time APIs

- String, misc and style changes

- Majority of work will be done for 2.10
Upstream work on OpenSFS/Intel branch

• Completion of UAPI header work for 2.10 release
  – LU-6401: make lustre header UAPI compliant
  – LU-6245: make libcfs/LNet headers UAPI compliant
    • No more libcfs or kernel headers in user land.

• Latest Infiniband support
  – LU-8874: Infiniband API has changed greatly in newer kernels. Change going to both master and upstream
    – Greatly improved MLX5 support

• Move to sphinx document API (LU-8919) (2.11)
• Don’t use linux linked list in user land (2.11)
• Tracepoint support (2.11)
What needs to be done to leave staging?

- Final checkpatch cleanups
- Continue syncing to OpenSFS tree
  - Lustre 2.10 support upstream is next milestone
- Merge in UAPI header cleanup
- Remove link linked list use for kernel <-> user space interfaces
  - Lnet selftest and nodemap (not upstream)
- IOCTL redirect and general IOCTL cleanup
- Migrate debugging to tracepoint
- Changes to lustre based on kernel maintainer reviews
Lustre server future in the Linux kernel

• All code improvements for clients are applied to servers
• LU-20 : Goal of no more patching the Lustre server’s kernel.
  – LU-684 : use dm_flakey for fail over testing
  – Module osd-ldiskfs can function with both patches and unpatched kernels without rebuild!!!
  – Almost there. Could drop rest of the patches.
  – Packaging is cleaned up.
• LU-6220 : Push most ldiskfs patches upstream
• Support ldiskfs up to 4.4 kernels
Conclusion

• Another successful year
• Major changes missing in master from upstream almost done being ported
• Convergence of OpenSFS/Intel and upstream client is almost complete
• Work for UAPI header cleanup after 5 years is finishing up.
• Upstream client at point people are actually using it
• Upstream client is starting to undergo code review from kernel maintainers